Машинное обучение для решения исследовательских и инженерных задач в науках о Земле

Михаил Криницкий

к.т.н., н.с.
Институт океанологии РАН им. П.П. Ширшова

Лаборатория взаимодействия океана и атмосферы и мониторинга климатических изменений (ЛВОАМКИ)

ПЛАН ЛЕКЦИИ

- ИНС: повторение теории
- ИНС: практика

ANN
ANN

Линейная регрессия

$$y_i \sim \mathcal{N}(\mu(\theta, x_i), \sigma^2)$$

$$\eta_i = \theta \cdot x_i$$

$$\mu(\theta, x_i) = \boldsymbol{\eta_i}$$

$g^{-1} = I$

Логистическая регрессия

$$y_i \sim \mathcal{B}\big(p(\theta, x_i)\big)$$

$$p(\theta, x_i) \propto \exp(\theta_1 \cdot x_i)$$

$$\eta_i^1 = \theta_1 \cdot x_i = \ln \frac{p_i}{1 - p_i}$$

$g^{-1} = \text{sigmoid}$

$$p(\theta_1, x_i) = \mathbf{sigmoid}(\boldsymbol{\eta_i^1}) = \frac{1}{1 + \exp(-\eta_i^1)}$$

Мультиномиальная логистическая регрессия

$$y_{ik} \sim \mathcal{B}\big(p_k(\theta, x)\big)$$

$$p_k(\theta, x_i) \propto \exp(\theta_k \cdot x_i)$$

$$\eta_i^k = \theta_k \cdot x_i = \ln p_{ik} + C$$

$g^{-1} = \text{softmax}$

$$p_k(\theta_1, x_i) = \mathbf{softmax}\big(\boldsymbol{\eta_i^k}, \{\boldsymbol{\eta_i^k}\}\big) = \frac{\exp \eta_i^k}{C^*}$$

Альтернативно (некорректно, но легче понять):

Обобщенные линейные модели: модели, в которых мат.ожидание параметра распределения целевой переменной вычисляется как некоторая функция $\boldsymbol{g^{-1}}(\cdot)$ от линейной функции η признакового описания объектов (событий) x_i

$\boldsymbol{g}(\cdot)$ – т.н. функция связи (link function)

Регрессия

$$y_i \sim \mathcal{N}(\mu(\theta, x_i), \sigma^2)$$

$$\eta_i = \theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right]$$

$$\mu(\theta, x_i) = \boldsymbol{\eta_i}$$

Бинарная классификация

$$y_i \sim \mathcal{B}\big(p(\theta, x_i)\big)$$

$$p(\theta, x_i) \propto \exp(\theta_1 \cdot x_i)$$

$$\eta_i^1 = \theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right] = \ln \frac{p_i}{1 - p_i}$$

$$p(\theta_1, x_i) = \mathbf{sigmoid}(\boldsymbol{\eta_i^1}) = \frac{1}{1 + \exp(-\eta_i^1)}$$

Мультиномиальная классификация

$$y_{ik} \sim \mathcal{B}\big(p_k(\theta, x)\big)$$

$$p_k(\theta, x_i) \propto \exp(\theta_k \cdot x_i)$$

$$\eta_i^k = \theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right] = \ln p_{ik} + C$$

$$p(\theta_1, x_i) = \mathbf{softmax}(\boldsymbol{\eta_i^k}, \{\boldsymbol{\eta_i^k}\}) = \frac{\exp \eta_i^k}{C^*}$$

Альтернативно (некорректно, но легче понять):

Обобщенные аддитивные модели – это обобщенные линейные модели, в которых в качестве признаков объектов x_i выступают некоторые гладкие (часто нелинейные) функции этих признаков

Как зависит вид функции потерь от вида функции связи?

НИКАК

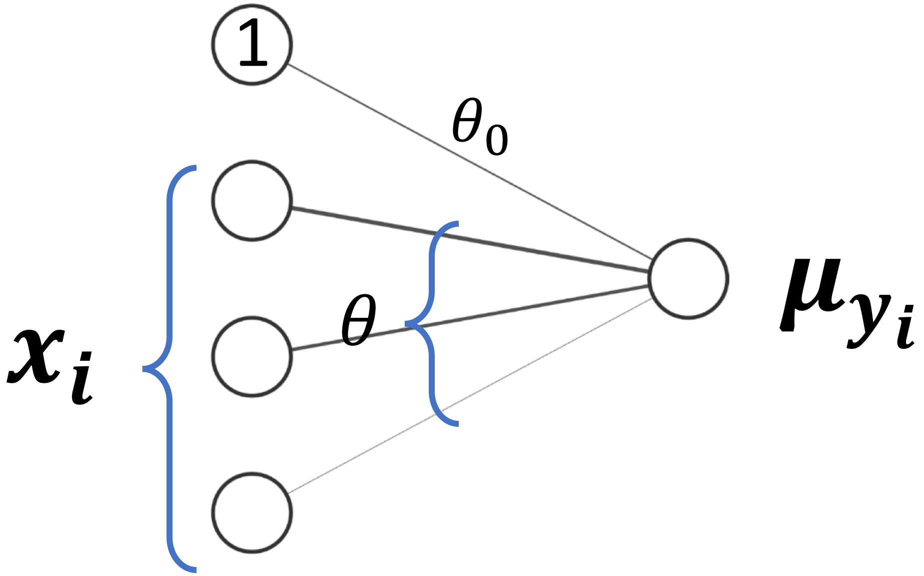
вид функции потерь $\mathcal{L}(\mathcal{T},\theta)$ зависит от вида распределения целевой переменной y_i, но не от вида функции связи $g(\cdot)$

- обратная функция связи $g^{-1}(\cdot)$ должна отображать R^1 (множество значений произвольной линейной функции $\theta \cdot x_i$) на множество параметров распределения переменной y_i
- вычисление параметра (параметров) распределения переменной y_i в модели производится согласно принципу GLM
- при таких условиях вычисление правдоподобия выборки $\mathcal{T}$ производится независимо от вида функции связи $g(\cdot)$ => вычисление функции потерь в подходе максимизации правдоподобия также производится независимо от вида функции связи $g(\cdot)$

previously on ML4ES

Почти всегда в случае задачи регрессии: $\boldsymbol{y} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu},\ \boldsymbol{\sigma}^2)$

ЛР: $\mu_{y_i} = \theta \cdot x_i + \theta_0$



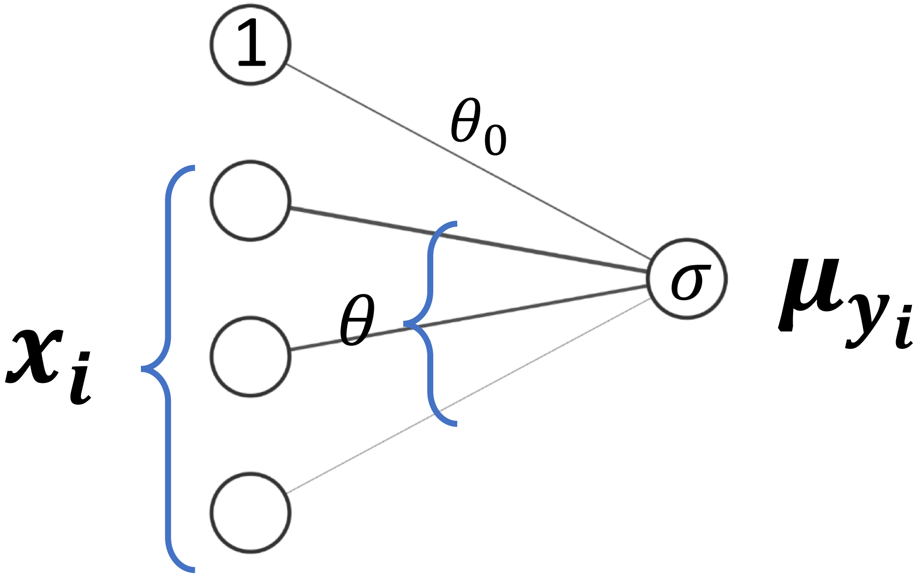
previously on ML4ES

Почти всегда в случае задачи регрессии: $\boldsymbol{y} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu},\ \boldsymbol{\sigma}^2)$

ЛР: $\mu_{y_i} = \theta \cdot x_i + \theta_0$

GLM: $\mu_{y_i} = \sigma(\theta \cdot x_i + \theta_0)$

$g^{-1} = \sigma$



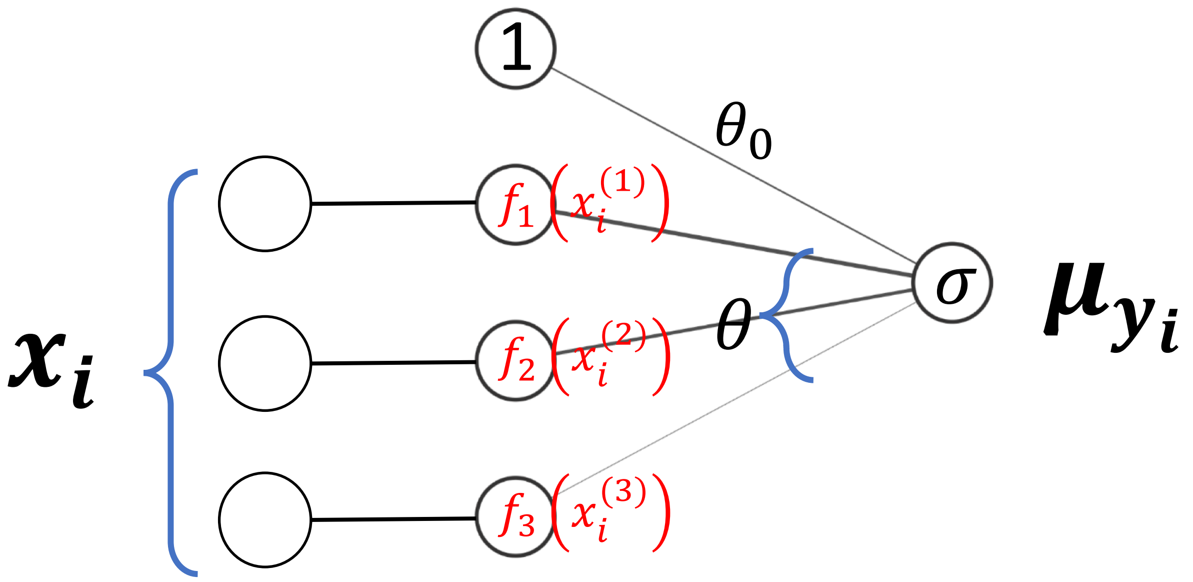
previously on ML4ES

Почти всегда в случае задачи регрессии: $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu},\ \boldsymbol{\sigma}^2)$

ЛР: $$\mu_{y_i} = \theta \cdot x_i + \theta_0$$

GLM: $$\mu_{y_i} = \sigma(\theta \cdot x_i + \theta_0)$$

GAM: $$\mu_{y_i} = \sigma\left(\theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right] + \theta_0\right)$$



previously on ML4ES

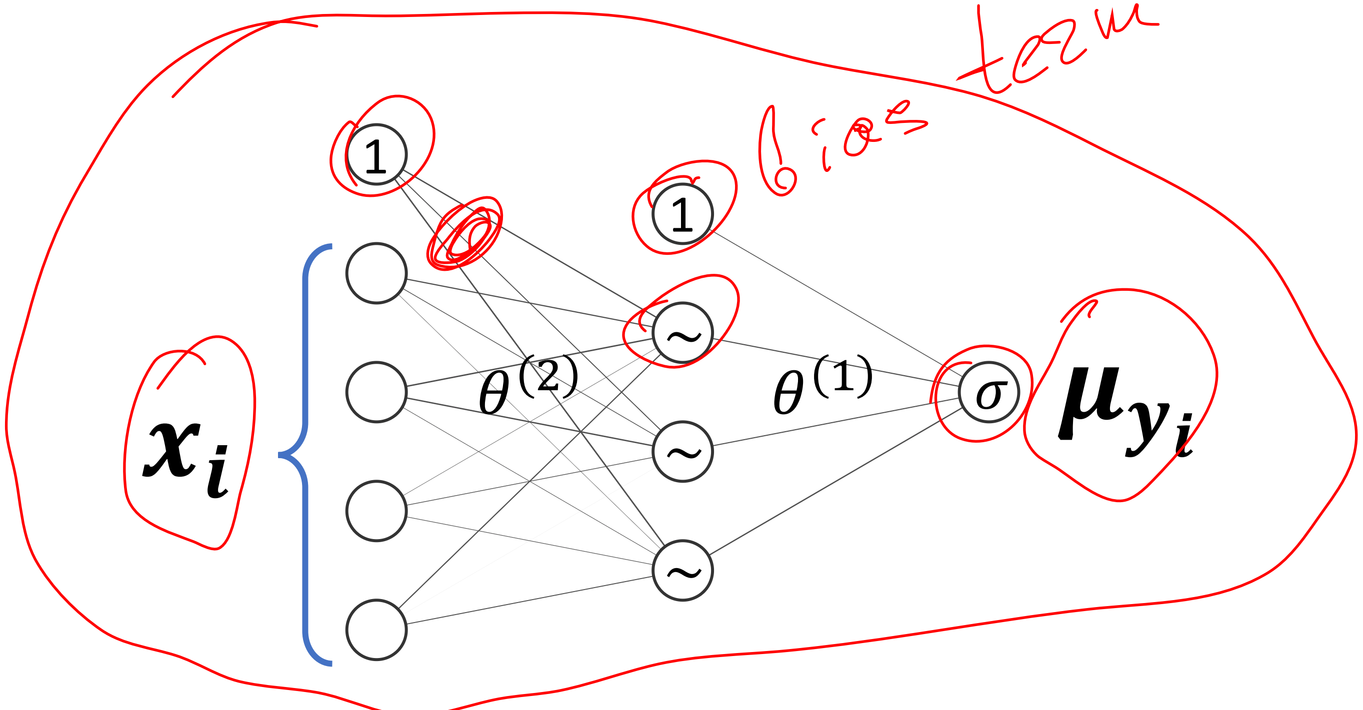
Почти всегда в случае задачи регрессии: $\boldsymbol{y} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu},\ \boldsymbol{\sigma}^2)$

ЛР: $\mu_{y_i} = \theta \cdot x_i + \theta_0$

GLM: $\mu_{y_i} = \phi(\theta \cdot x_i + \theta_0)$

GAM: $\mu_{y_i} = \sigma\left(\theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right] + \theta_0\right)$

ИНС: $\mu_{y_i} = \phi\left(\theta_0^{(2)} + \theta^{(2)} \cdot \phi\left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i\right)\right)$



previously on ML4ES

Почти всегда в случае задачи регрессии: $\boldsymbol{y} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu},\ \boldsymbol{\sigma}^2)$

ЛР: $\mu_{y_i} = \theta \cdot x_i + \theta_0$

GLM: $\mu_{y_i} = \phi(\theta \cdot x_i + \theta_0)$

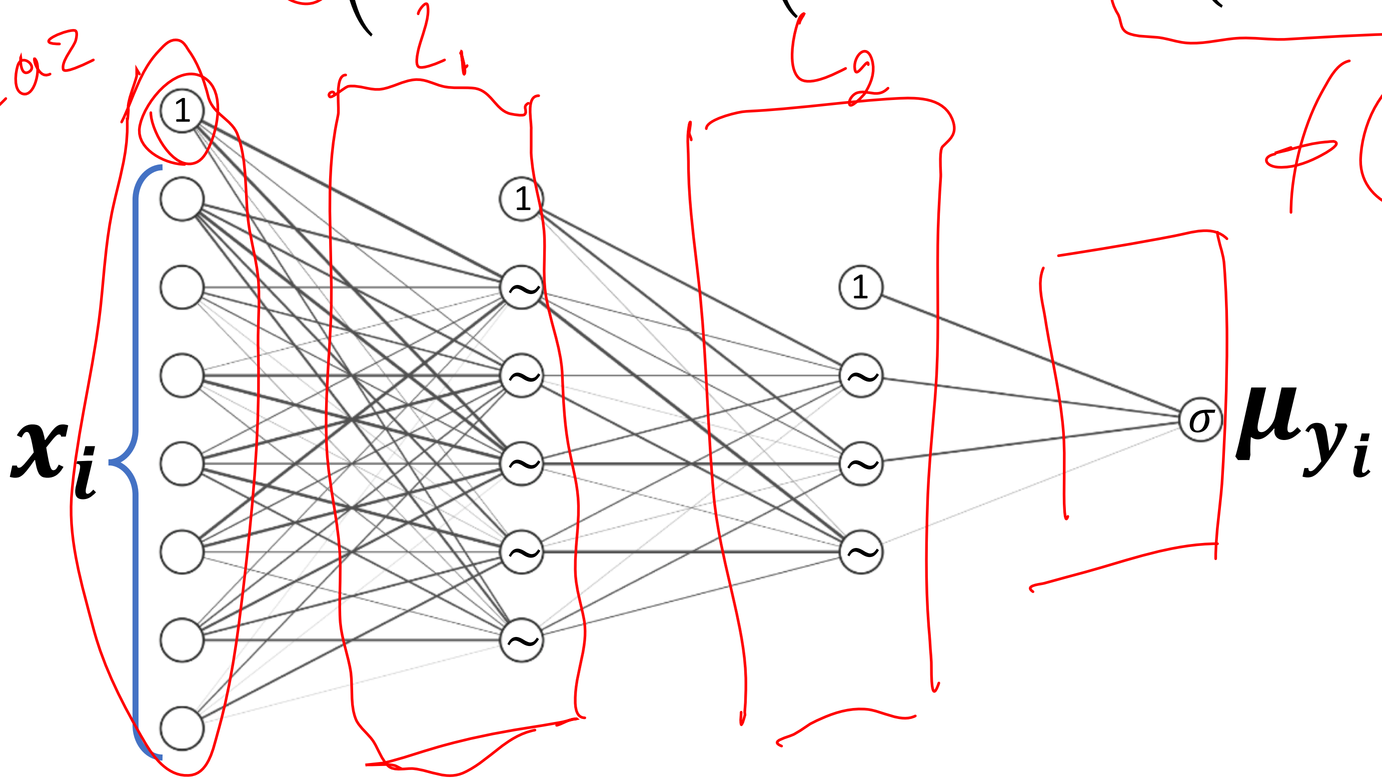
GAM: $\mu_{y_i} = \sigma\left(\theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right] + \theta_0\right)$

ИНС: $\mu_{y_i} = \phi\left(\theta_0^{(3)} + \theta^{(3)} \cdot \phi\left(\theta_0^{(2)} + \theta^{(2)} \cdot \phi\left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i\right)\right)\right)$

$f(\theta^{(1)} x_i)$

ReLU
rectified linear unit

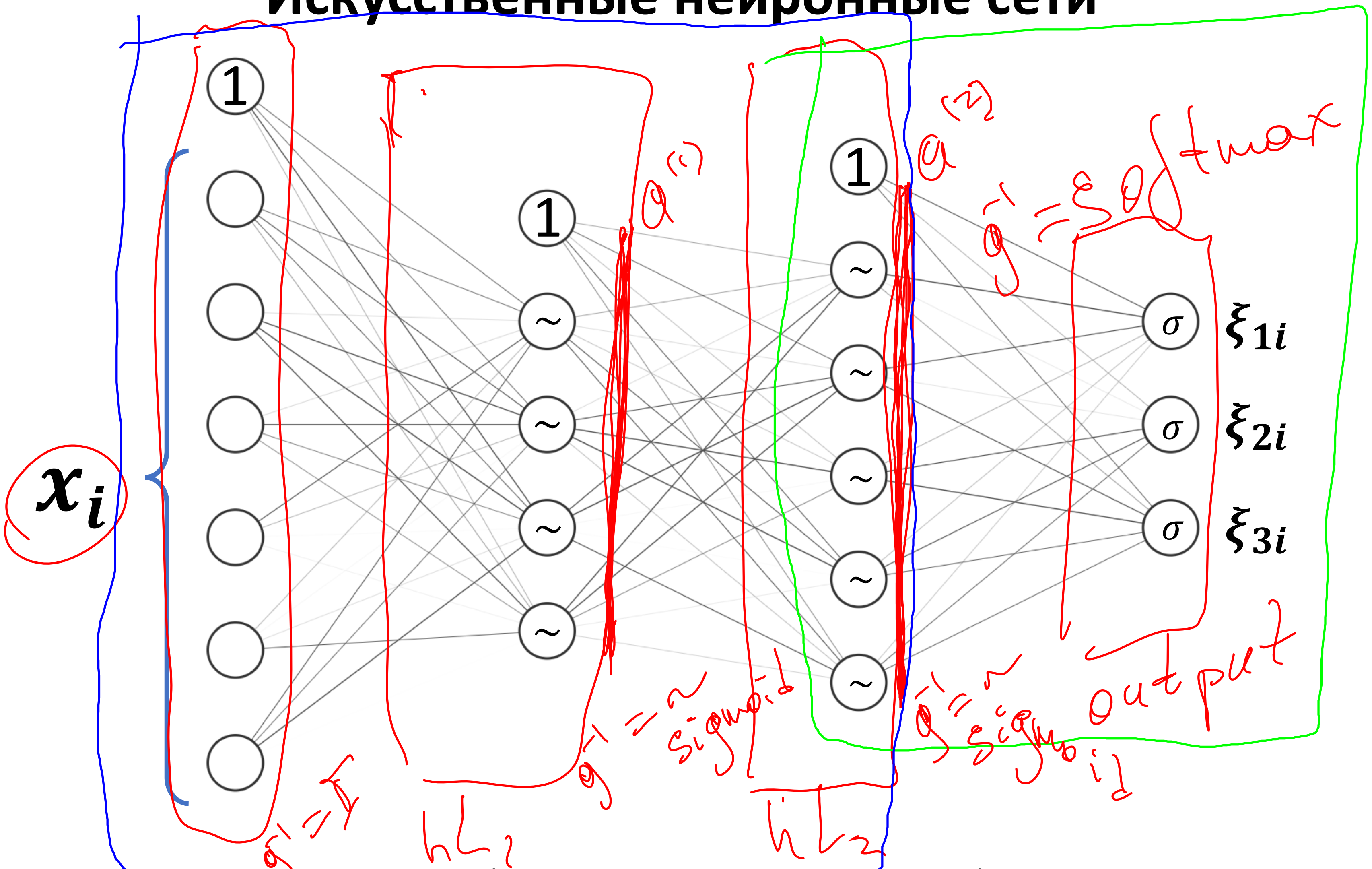
softmax
sigmoid



L_1 L_2

– скрытые слои
hidden layers

Искусственные нейронные сети

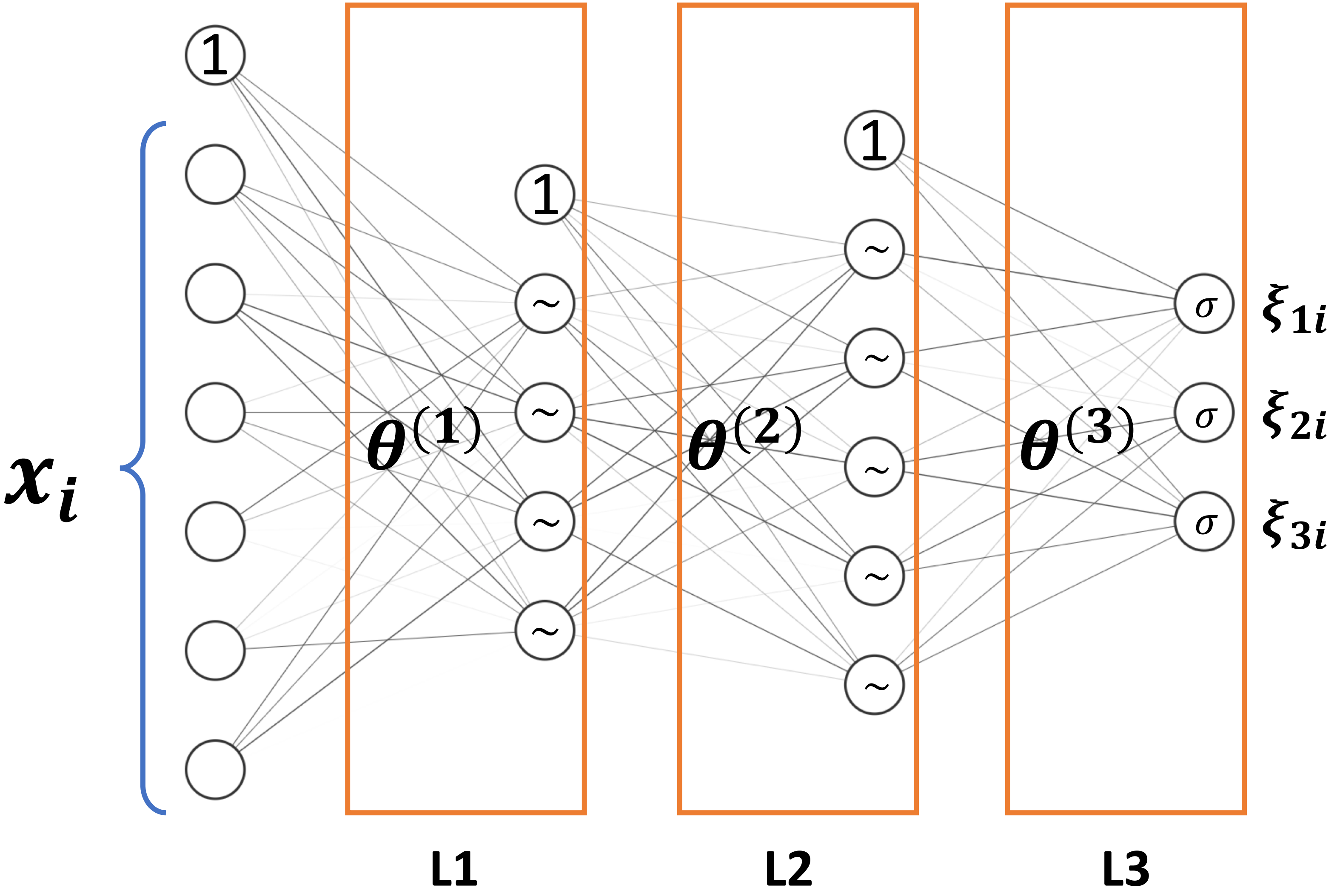


Вид ИНС: многослойный перцептрон (multilayer perceptron, MLP)
Feedforward NN (FNN, сеть прямого распространения)
полносвязная ИНС (Fully-connected NN, FCNN)

previously on ML4ES

MLP

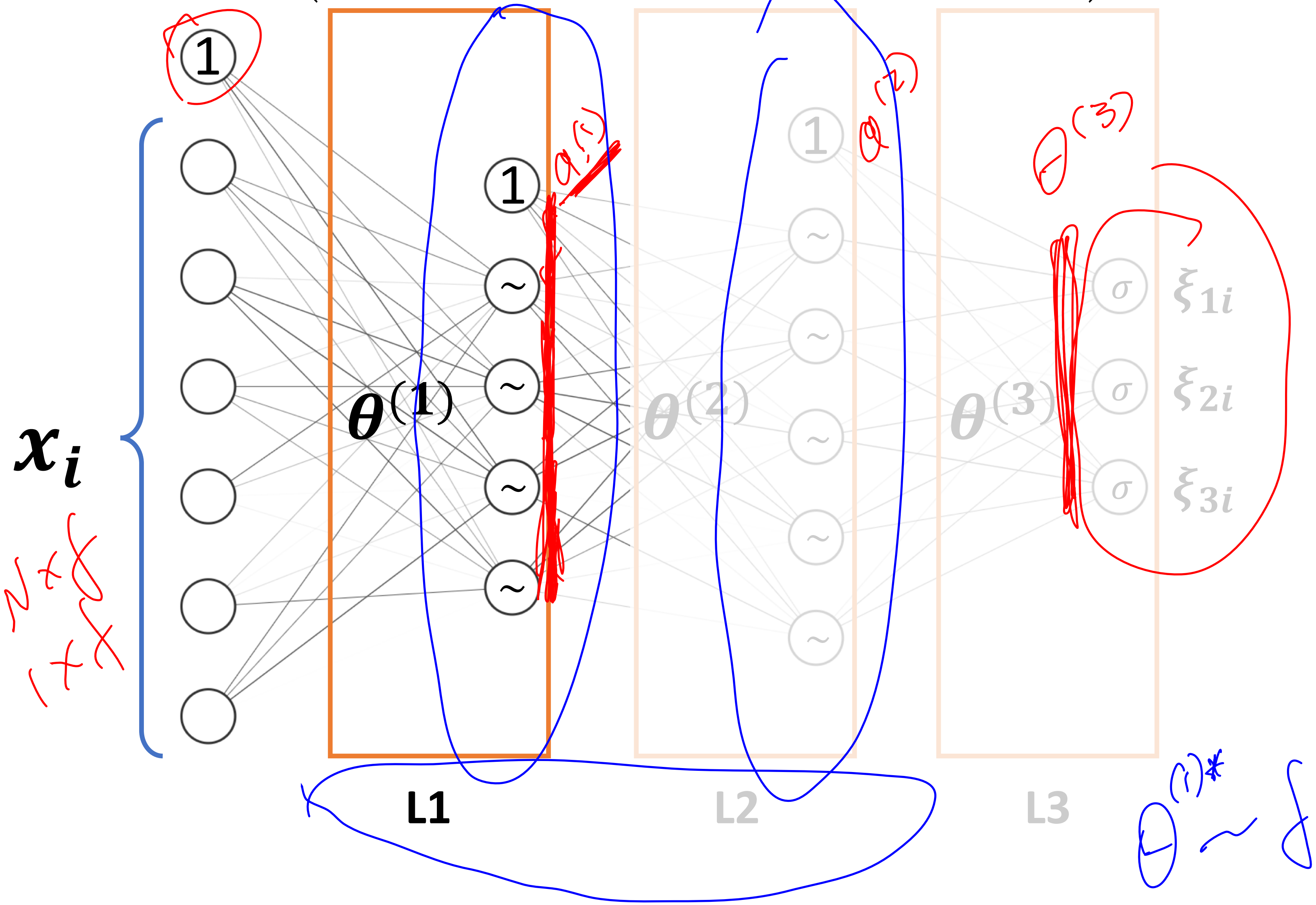
$$\xi_i = \sigma\left(\theta_0^{(3)} + \theta^{(3)} \cdot \phi\left(\theta_0^{(2)} + \theta^{(2)} \cdot \phi\left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i\right)\right)\right)$$



previously on ML4ES

MLP

$$\xi_i = \sigma\left(\theta_0^{(3)} + \theta^{(3)} \cdot \phi\left(\theta_0^{(2)} + \theta^{(2)} \cdot \boxed{\phi\left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i\right)}\right)\right)$$



$z^{(1)} = X^* \cdot \theta^{(1)*}$

$a^{(1)} = \phi(z^{(1)})$

$X \quad 1 \times f$

$z \sim 1 \times f \cdot f \times c^{(1)}$

$z \sim 1 \times c^{(1)}$

$a \sim \phi(z) \sim z$

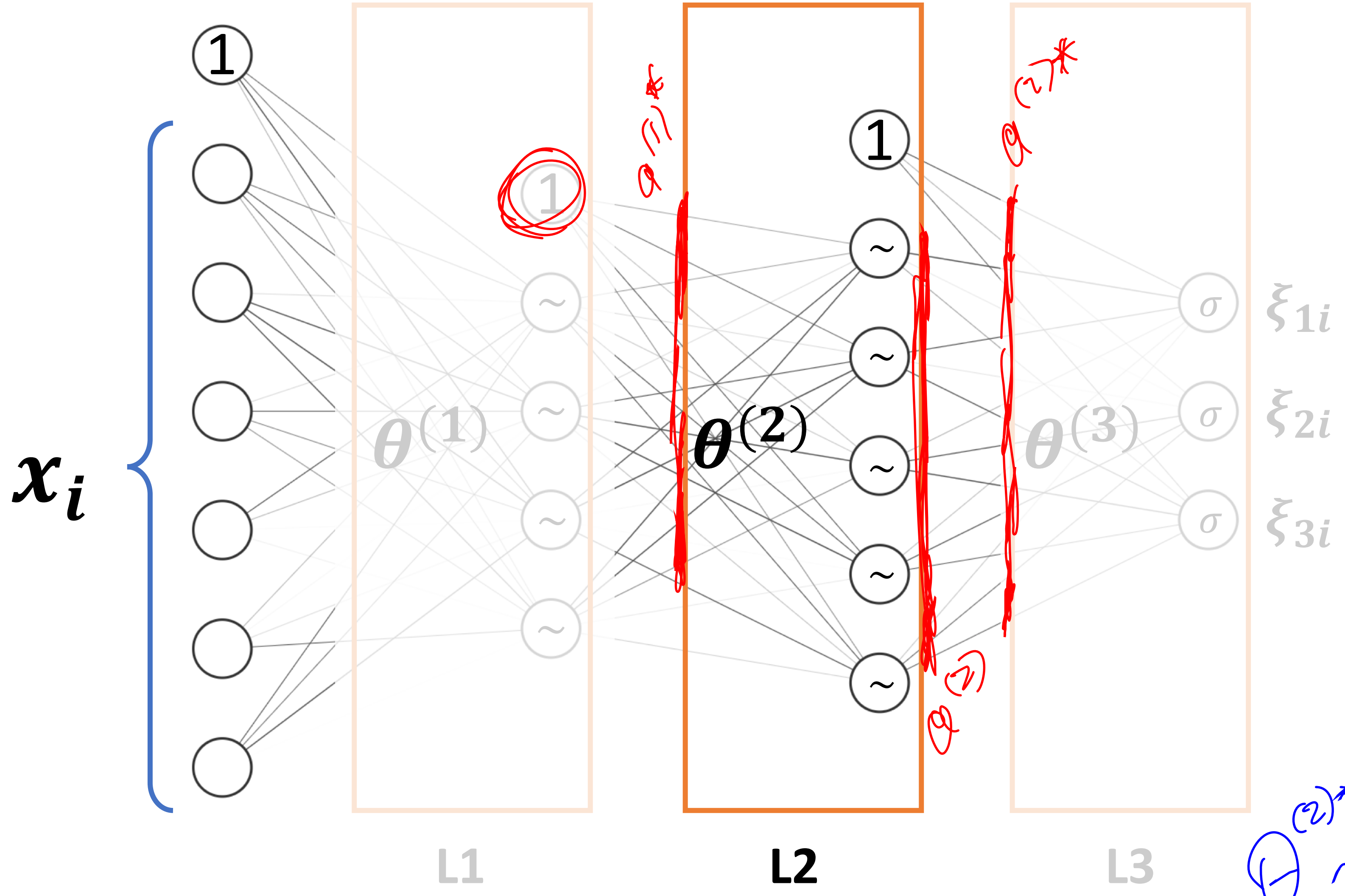
$\theta^{(1)*} \sim f \times c^{(1)} \quad a \sim 1 \times c^{(1)}$

10

previously on ML4ES

MLP

$$\xi_i = \sigma\left(\theta_0^{(3)} + \theta^{(3)} \cdot \boxed{\phi\left(\theta_0^{(2)} + \theta^{(2)} \cdot \phi\left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i\right)\right)}\right)$$



$c^{(1)} = 10$

$\dfrac{(\times c^{(1)})}{a^{(1)}} \rightarrow a^{(1)*}$ $\times (c^{(1)}+1)$

$z^{(2)} = a^{(1)*} \cdot \theta^{(2)*}$

$a^{(2)} = f(z^{(2)})$

$c^{(2)} = 8$

$\theta^{(2)*} \sim (c^{(1)}+1) \times c^{(2)}$

$z^{(2)} \sim N \times c^{(2)}$

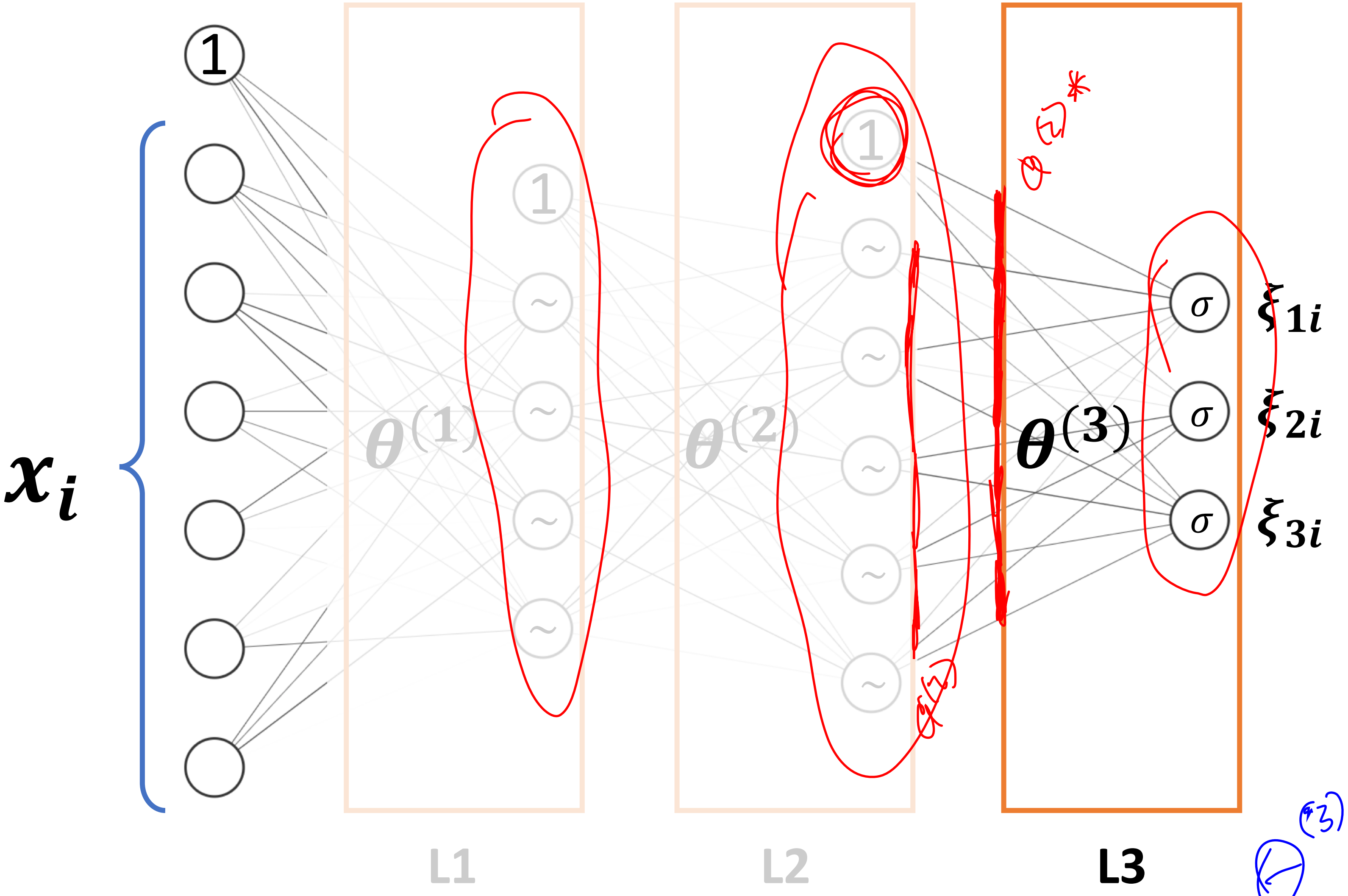
$a^{(2)} \sim N \times c^{(2)}$

$\theta^{(2)*} \sim (c^{(1)}+1) \times c^{(2)}$

previously on ML4ES

MLP

$$\xi_i = \boxed{\sigma\left(\theta_0^{(3)} + \theta^{(3)} \cdot \phi\left(\theta_0^{(2)} + \theta^{(2)} \cdot \phi\left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i\right)\right)\right)}$$



$$\theta^{(2)*} \sim N \times (c^{(2)}+1)$$

$$z^{(3)} = \theta^{(2)*} \cdot \theta^{(3)*}$$

$$z^{(3)} \sim N \times (c^{(2)}+1) \quad (c^{(2)}+1) \times c^{(3)}$$

$$\xi = \sigma(z^{(3)})$$

$$\xi \sim z^{(3)}$$

$$z^{(3)} \sim N \times c^{(3)} \quad 3$$

$$\theta^{(3)*} \sim (c^{(2)}+1) \times c^{(3)}$$